

Depositing Electron Microscopy Maps

Meeting Review

Stephen D. Fuller*
Division of Structural Biology
Wellcome Trust Centre for Human Genetics
University of Oxford
Roosevelt Drive
Oxford OX3 7BN
United Kingdom

A meeting was held at the European Bioinformatics Institute (EBI) in Hinxton, United Kingdom to discuss recent progress in the development of EMD, a database for maps determined by electron microscopy that is now integrated with MSD, the macromolecular structure database at EBI. This meeting of representatives of many of the major image processing groups in electron microscopy also discussed possible software developments that would ease the documentation and deposition of such datasets. The meeting concluded with a strong endorsement of map deposition in electron microscopy and its linkage with the family of archival databases in biomedical research.

The maturation of electron microscopy as a structural technique over the past decade has been spectacular. Much of the recent progress results from two complementary advances. The first was the introduction of cryo-electron microscopy (cEM), a technique that preserves native structures in vitrified water. The second was the development of image processing techniques that allowed the determination of three-dimensional structure from the relatively low-contrast image generated by cryo-electron micrographs of unstained specimens. These advances make electron microscopy an important complement to NMR and X-ray diffraction. An increasing number of projects use this complementarity directly. Often a three-dimensional structure derived from cEM of a large complex is used to generate an atomic model in combination with structures of components determined by X-ray crystallography or NMR. The transfer of information is not only from high resolution to low. Some X-ray crystallographic structures have been solved by using models derived from microscopy to determine initial phases.

This recent progress in microscopy as a structural technique shows the power and desirability of sharing the information from microscopy within the structural community. This was not routinely possible until recently. This situation has changed as a result of coordinated efforts by the Protein Data Bank (PDB) at RCSB and the Macromolecular Structure Database (MSD) at EBI. An X-ray crystallographic or NMR experiment produces atomic coordinates that are routinely deposited along with supporting data. A cEM experiment produces a map showing the electron potential distribution within

the specimen. A few cEM studies produce atomic models, however, many, particularly at lower resolution, only produce maps. Deposition in the PDB requires coordinates. Depositing the map was possible by the indirect route of depositing structure factors calculated from the map along with the coordinates of the atomic model. A number of cEM structures have been deposited this way. These represent an important step, as they allow direct comparison with other techniques. A specific EM-oriented template has been developed to simplify and standardize these depositions. Most EM structures lack the coordinates needed for PDB deposition. Only a handful of the hundred or so EM structures determined each year are available in the PDB. This situation limits the usefulness of microscopy as a structural technique.

A barrier to the routine deposition of EM structures is the lack of agreement on an image or map format. There are at least a dozen image and file formats in common use in EM image processing and little agreement on standardization. A meeting was held at the European Bioinformatics Institute (EBI) near Cambridge to try to explore solutions to this problem and others that limit the deposition of EM structures. The Wellcome Trust-supported meeting was held under the auspices of the Information Interchange for Molecular Structures (IIMS) initiative of the European Union. It assembled representatives of many of the active labs in software development for electron microscopy in Europe and the United States. This workshop was also an opportunity to acquaint the community with the progress that has been made in establishing a database for electron microscopy maps. Although the meeting was held in Cambridge, the ongoing cooperation between the EBI and the RCSB will guarantee that their efforts will continue to be coordinated.

The current EM database (EMD) profits from the experience with Biolumage, a previous Commission-funded project that established a limited collection of light and electron microscopy datasets. The development of the EM database at EBI was funded under the Fifth Framework Program of the European Commission. Deposition interface tools have been deployed by both EBI and RCSB to capture coordinate data derived from electron microscopy experiments using the data definitions developed in cooperation between the two institutions. Map and volume data are collected by the EBI and converted to the format of the Collaborative Computing Project Number 4 (CCP4), the most common map format used by protein crystallographers. This choice is particularly appropriate because one of the aims of the database is to enhance the interaction between microscopy and other structural techniques. The database of maps and volumes is linked to the MSD at EBI and the RCSB/PDB. The map and volume data collected by IIMS are treated in the same archival fashion as PDB data and are explicitly linked to coordinate models derived from these data.

The data provided would be far more valuable if they were accompanied by criteria of reliability that would

*Correspondence: stephen@strubi.ox.ac.uk

aid the interpreter. There is not a single common standard for resolution and reliability that covers the entire range of possible entries. There are five types of data that form the bulk of electron microscopy structures: electron crystallographic, symmetric single particles such as icosahedral, helical structures, single particles without symmetry, and tomography. The criteria for reliability in each of these techniques vary. Crystallographic and helical reconstructions can use the measured spatial resolution of their data and such indicators as phase error as criteria. The reproducibility of the reconstruction, as assessed by the plot of the Fourier Shell correlation (FSC) between two independent reconstructions, is often used for single particle reconstruction. The participants in the workshop agreed that a single resolution number would be less useful than the curve showing the FSC as a function of spatial frequency. It was agreed that this curve should be deposited with single particle reconstructions along with any mask used for its calculation. Tomographic reconstructions are derived from multiple tilted views of a single object and hence the fineness of the sampling is used as a criterion for reliability and resolution. New measures of reliability should be developed as a larger number of structures are deposited.

The workshop discussed many of the issues that arise from deposition and the presentation of the data in a form that would be useful to nonspecialists. It was agreed to begin to develop "harvest files," similar to those currently used in X-ray crystallography and NMR. The common image processing packages would be modified to automatically assemble the important information from image processing and imaging and generate these files. This will ease the deposition of the documented final map.

Initially, the database will implement immediate release of the header information characterizing the entries but provide hold periods ranging from immediate release to four years for the maps and structure factors. This may seem at odds with the now standard policy within crystallography of relatively rapid release. This reticence is perhaps easier to understand when one remembers that microscopy is still a relatively new structural technique. As the drive for high-throughput methods spreads through microscopy, one can expect the speed of release to increase as it has for other structural techniques. This first step is important because it establishes the database as a resource for the community. The participants in the workshop noted this by agreeing to the following statement:

We note that the European Bioinformatics Institute (EBI) through the Macromolecular Structure Database (MSD) now provides a permanent resource for the deposition of three-dimensional maps derived by electron microscopy (see www.ebi.ac.uk/msd-srv/emdep). In addition, coordinate data derived from these maps are deposited in the PDB archive for macromolecular structural data. We intend to use these facilities for the routine deposition of maps and coordinate data produced by our work. These databases are open to the international community and will become part of a family of linked databases in biomedical research.

We encourage our colleagues to follow our example by submitting maps from their work, at the stage of publication, to these archival databases.

Acknowledgments

The author thanks Dr. Kim Henrick and Dr. Richard Newman (EMBL-EBI, Hinxton, UK), Dr. José-Maria Carazo (CNB-CSIC, Madrid, Spain), Dr. Rishi Matadeen (University of Oxford, Oxford, UK), and Dr. Helen Berman (RCSB, Rutgers University, New Jersey) for their comments on this report. A list of participants and documentation from the meeting can be found at www.ebi.ac.uk/msd/MSDProjects/IIMSReport.html.